

Improved approximation algorithms and disjunctive relaxations for some knapsack problems

Yuri Faenza*, Igor Malinović†

January 26, 2017

Abstract

We consider two knapsack problems. The time-Invariant Incremental Knapsack problem (IIK) is a generalization of Maximum Knapsack to a discrete multi-period setting. At each time the capacity increases and items can be added, but not removed from the knapsack. The goal is to maximize the sum of profits over all times. While IIK is strongly NP-Hard [8], we design a PTAS for it and some of its generalizations.

The Minimum Knapsack problem (MIN-K) aims at minimizing a linear function over the 0/1 points that satisfy a single linear constraint. Despite the existence of an FPTAS, it is an open question whether one can obtain a poly-size linear formulation with constant integrality gap for MIN-K. This motivated recent work on disjunctive formulations having integrality gap of at most $(1 + \varepsilon)$ for a fixed objective function. We give such a formulation of size polynomial in n and subexponential in $\frac{1}{\varepsilon}$.

1 Introduction

Knapsack problems are among the most fundamental and most studied in integer programming. Some variants forego the development of modern combinatorial optimization, dating back to at least 1896 [21]. The best known representative of this class is arguably *Maximum Knapsack* (MAX-K): given a set of items, each having a profit and a weight, and a threshold capacity, find a most profitable subset of items whose total weight does not exceed the threshold. MAX-K is known to be NP-complete [16], while admitting a *fully polynomial-time approximation scheme* (FPTAS) [14]. Many classical algorithmic techniques including greedy, dynamic programming, backtracking/branch-and-bound have been studied by means of solving this problem, see e.g. [17]. The algorithm of Martello and Toth [20] has been known to be the fastest in practice for exactly solving knapsack instances [1].

Resembling real-world scenarios, many recent works studied extensions of classical combinatorial optimization problems to multi-period settings, see e.g. [13], [24], [25]. Bienstock et al. [8] proposed an interesting generalization of MAX-K that they dubbed *time-Invariant Incremental Knapsack* (IIK). In IIK, we are given a set of items $[n]$ with profits $p : [n] \rightarrow \mathbb{R}_{>0}$ and weights $w : [n] \rightarrow \mathbb{R}_{>0}$ and a knapsack with non decreasing capacity b_t over time $t \in [T]$, i.e. $0 < b_t \leq b_{t+1}$ for $t \in [T - 1]$. We are allowed to add items at each time as long as the capacity constraint is not violated, and once inserted, an item cannot be removed from the knapsack. The goal is to maximize the total profit, which is defined to be the sum, over $t \in [T]$, of profits of items in the knapsack at time t . IIK models a scenario where available resources (e.g. money, labour force) augment over time in a predictable way, allowing an increase of our portfolio. Take as an example a bond market with an extremely low levels of volatility, where all coupons render profit only at their common

*IEOR, Columbia University, USA. Email: yf2414@columbia.edu

†DISOPT, EPFL, Switzerland. Email: igor.malinovic@epfl.ch

maturity time T (*zero-coupon* bonds) and an increasing budget over time that allows buying more and more (differently sized and priced) packages of those bonds. For variations of MAX-K that have been historically used to model financial problems, see [17]. IIK was proved in [8] to be strongly NP-hard. In the same paper, a PTAS is given when $T = O(\log n)$. This result is stronger than the one from [24], where a PTAS for the special case $p = w$ is given when T is a constant. Again when $p = w$, a $1/2$ -approximation algorithm for generic T is given in [13]. Results from [26] can be adapted to give an algorithm that solves IIK in time polynomial in n and of order $(\log T)^{O(\log T)}$ for a fixed approximation guarantee ε [23]. However, nothing was known on the approximability of the problem for generic T and profit/weight structure.

Pure knapsack problems are rarely studied independently in either theoretical or practical applications. One often aims at developing techniques that remain valid when more general, less structured constraints are added. This can be achieved by casting the problem in a geometric setting, mostly using *linear* and *semidefinite* programming (LP and SDP, respectively). Moreover, LP-based techniques have been broadly used for designing approximation algorithms for knapsack-related problems, including multiple knapsack, generalized assignment, and scheduling problems among others, see e.g. [9], [12], [15]. The standard LP relaxation for MAX-K has integrality gap (IG) 2, but one can design an (extended) LP relaxation with $(n + 1/\varepsilon)^{O(1/\varepsilon)}$ variables and constraints, and IG $1 + \varepsilon$ [6]. (Interestingly, no such relaxation exist in the original space [11]). Hence, the applicability of LP to MAX-K is essentially understood, the only (yet very interesting) open question being whether relaxations of size $\text{poly}(n, 1/\varepsilon)$ with IG $1 + \varepsilon$ exist.

Surprisingly, the situation is quite different for *Minimum Knapsack* (MIN-K). This is the minimization version of the classical MAX-K: given a set of n items with non-negative costs c and weights w , find a cheapest subset of the items whose total weight is at least the target value β . NP-completeness of MIN-K immediately follows from the NP-completeness of MAX-K, and it is not hard to show that the FPTAS for MAX-K can be adapted to MIN-K [14]. MIN-K is an important problem appearing as a substructure in many IPs. Valid inequalities for MIN-K – like *knapsack cover inequalities* – have been generalized e.g. to problems in scheduling and facility location. Unlike MAX-K, there cannot be an LP relaxation in the original space with constant integrality gap and a polynomial number of inequalities [10]. More generally, the problem seems to be very challenging for geometric algorithms, as even the Lasserre strengthening of the natural LP relaxation may have unbounded integrality gap up to level $n - 1$ [19]. No relaxation for the problem with a constant integrality gap and a polynomial number of inequalities is known. A remarkable recent result [4] shows that for any fixed $\varepsilon > 0$ there is an LP relaxation of size $n^{O(\log n)}$ with IG bounded by $2 + \varepsilon$. Clearly the problem becomes easier if we ask for a relaxation of MIN-K that has constant integrality gap *only* for a fixed objective function. Indeed, let x' be a $(1 + \varepsilon)$ -approximate solution obtained using an FPTAS. Adding $c^T x \geq c^T x' / (1 + \varepsilon)$ to the standard LP relaxation reduces the integrality gap to $1 + \varepsilon$. A more interesting question is requiring the relaxation to be more structured, in order to be able to reuse it to tackle a broader class of knapsack problems. A common technique to produce those structured relaxations is *Disjunctive programming*. It provides a very general and flexible approach to find strong relaxations for integral sets, and it has been exploited in practice to produce the so called *disjunctive cuts* for MILP. See Appendix B for more details on disjunctive programming and disjunctive cuts.

Several papers successfully applied disjunctive programming to obtain theoretical results on knapsack problems, including MIN-K [7] and MAX-K [6]. In particular, in [7], a linear relaxation for MIN-K with $n^2(\lceil 1/\varepsilon \rceil)^{\lceil C_\varepsilon \rceil}$ variables and constraints and integrality gap $1 + \varepsilon$ is given, where $C_\varepsilon = \log_{1+\varepsilon} \frac{1}{\varepsilon} = \Theta(\frac{1}{\varepsilon} \log(\frac{1}{\varepsilon}))$. Hence, the relaxation has size polynomial in n and exponential in $\frac{1}{\varepsilon}$.

Our Contributions. The first main result of this paper is an algorithm for computing a $(1 - \varepsilon)$ -approximated solution for IIK that depends polynomially on the number n of items and, for any fixed ε , also polynomially on the number of times T . In particular, our algorithm provides a PTAS for IIK, regardless of T .

Theorem 1. *There exists an algorithm that, when given as input $\varepsilon \in \mathbb{R}_{>0}$ and an instance \mathcal{I} of IIK with n items and T times, produces a $(1 - \varepsilon)$ -approximation to the optimum solution of \mathcal{I} in time $O(T^{h(\varepsilon)} \cdot n f_{LP}(n))$. Here $f_{LP}(m)$ is the time required to solve a linear program with $O(m)$ variables and constraints, and $h : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{\geq 1}$ is a function depending on ε only. In particular, there exists a PTAS for IIK.*

Theorem 1 therefore dominates all previous results on IIK [8, 13, 24, 26]. Interestingly, it is based on designing a disjunctive formulation – a tool mostly common among integer programmers and practitioners – and then rounding the solution to its linear relaxation with a greedy-like algorithm. Because of the hardness result from [8], Theorem 1 is essentially optimal. We see Theorem 1 as an important step towards the understanding of the complexity landscape of knapsack problems over time. Theorem 1 is proved in Section 2: see Section 2.1 for a sketch of the techniques we use and a detailed summary of Section 2. In Section 2.5, we show some extensions of Theorem 1 to more general problems.

The second main result of this paper is a disjunctive relaxation for MIN-K of size polynomial in n and subexponential in $1/\varepsilon$, hence asymptotically smaller than the one provided in [7]. Recall that $C_\varepsilon = \Theta(\frac{1}{\varepsilon} \log(\frac{1}{\varepsilon}))$.

Theorem 2. *Given $\varepsilon > 0$ and a fixed objective function, there is a disjunctive relaxation for MIN-K with $n^2(1/\varepsilon)^{O(\sqrt{C_\varepsilon})}$ variables and constraints such that the integrality gap is $1 + \varepsilon$.*

Theorem 2 is proved in Section 3. An overview of the techniques used is given in Section 3.1.

More related work. The authors in [8] show that IIK is strongly NP-hard, and provide an instance showing that the natural LP relaxation has an unbounded integrality gap. Furthermore, [8] discusses the relation between IIK and the generalized assignment problem (GAP), highlighting the differences between those problems. In particular, there does not seem to be a direct way to apply to IIK the $(1 - 1/e - \varepsilon)$ approximation algorithm [12] for GAP. A special case of GAP where an item has non-changing weight and profit over the set of bins is called the multiple knapsack problem (MKP). MKP is strongly NP-complete as well as IIK and has an LP-based efficient PTAS (EPTAS) [15]. There is a certain similarity between the scheme in [15] and the one we are going to present here, since they are both based on reducing the number of possible profit classes and knapsack capacities, and then guessing the most profitable items in each class. However, the way this is performed is very different. The key ingredient of the approximation schemes so far developed for MKP is a shifting trick. In rounding a fractional LP solution it redistributes and mixes together items from different buckets. Applying this technique to IIK would easily violate the monotonicity constraint, i.e. $x_{t,i} \leq x_{t+1,i}$ where $x_{t,i}$ indicates whether an item i is present in the knapsack at time t . This highlights a significant difference between the problems: the ordering of the bins is irrelevant for MKP while it is crucial for IIK.

2 A PTAS for IIK

We already defined IIK in the introduction. The following IP gives an equivalent, mathematical programming formulation.

$$\begin{aligned} \max \quad & \sum_{t \in [T]} p^T x_t \\ \text{s.t.} \quad & w^T x_t \leq b_t \quad \forall t \in [T] \\ & x_t \leq x_{t+1} \quad \forall t \in [T-1] \\ & x_t \in \{0, 1\}^n \quad \forall t \in [T]. \end{aligned} \tag{1}$$

Recall that by the definition of the problem $0 < b_t \leq b_{t+1}$ for $t \in [T-1]$. We also assume wlog that $1 = p_1 \geq p_2 \geq \dots \geq p_n$.

2.1 Overview of the proof technique

In order to illustrate the ideas behind Theorem 1, let us recall a possible PTAS for the classical MAX-K with capacity β , n items, profit and weight vector p and w respectively. Recall the greedy algorithm for knapsack:

1. Sort items so that $\frac{p_1}{w_1} \geq \frac{p_2}{w_2} \geq \dots \geq \frac{p_n}{w_n}$.
2. Let $\bar{x}_i = 1$ for $i = 1, \dots, \bar{i}$, where \bar{i} is the maximum integer s.t. $\sum_{1 \leq i \leq \bar{i}} w_i \leq \beta$.

It is well-known that $p^T \bar{x} \geq p^T x^* - \max_{i \geq \bar{i}+1} p_i$, where x^* is the optimum solution to the fractional relaxation. A PTAS for MAX-K can then be obtained as follows: guess a set S_0 of $\lceil \frac{1}{\varepsilon} \rceil$ items with $w(S_0) \leq \beta$ and consider the “residual” knapsack instance \mathcal{I} obtained removing items in S_0 and items ℓ with $p_\ell > \min_{i \in S_0} p_i$, and setting the capacity to $\beta - w(S_0)$. Apply the greedy algorithm to \mathcal{I} as to obtain solution S . Clearly $S_0 \cup S$ is a feasible solution to the original knapsack problem. The best solutions generated by all those guesses can be easily shown to be a $(1 - \varepsilon)$ -approximation to the original problem.

When trying to extend the algorithm above to our setting, we face two problems. First, we have multiple times, and a standard guessing over all times will clearly be exponential in T . Second, when inserting an item in the knapsack in a specific time, we are clearly imposing this decision on all time stamps that succeed it, and it is not clear a priori how to take it into account.

We solve this by proposing an algorithm that, in a sense, still follows the general scheme of the greedy algorithm sketched above: after some preprocessing, guess items and insertion times that give high profit, and then fill the remaining capacity with an LP-driven integral solution. In particular, we first show that by losing at most a 2ε fraction of the profit we can assume the following (see Section 2.2): item 1, which has the maximum profit, is always inserted in the knapsack at some time; the capacity of the knapsack only increases and hence the insertion of items can only happen at $J = O(\frac{1}{\varepsilon} \log T)$ times (we call them *significant*); and the profit of each item is either much smaller than $p_1 = 1$ or it takes one of $K = O(\frac{1}{\varepsilon} \log \frac{T}{\varepsilon})$ possible values (we call them *profit classes*). This will give a 2-dimensional grid of size $J \times K$ of “significant times” vs “profit classes” with $O(\frac{1}{\varepsilon^2} \log^2 \frac{T}{\varepsilon})$ entries. Note that those entries are still too many to perform a guessing over all of them. Instead, we proceed as follows: for a carefully guessed subset of points (j, k) of this grid, we will either exactly guess how many items from profit class k are inserted at time j , or impose that they are at most $\frac{1}{\varepsilon}$. To each of those guesses, we associated a natural IP (see Section 2.3). The optimal solution x^* of its linear relaxation is not as simple as the classical fractional greedy solution, but it still has a lot of structure. We exploit this to produce an integral solution to the IP, and show that we can round x^* such that the portion of the profit we lose is negligible (see Section 2.4).

2.2 Reducing IIK to special instances and solutions

Our first step will be to show that we can reduce IIK, without loss of generality, to solutions and instances with a special structure. The first reduction is immediate: we restrict to solutions where the highest profit item is inserted in the knapsack at some time. We call these *1-in solutions*. This can be assumed by guessing which is the highest profit item that is inserted in the knapsack, and reducing to the instance where all higher profit items have been excluded. Since we have n possible guesses, the running time is scaled by a factor $O(n)$.

Observation 1. *Suppose there exists a function $f : \mathbb{N} \times \mathbb{N} \times \mathbb{R}_{>0}$ such that, for each $n, T \in \mathbb{N}$, $\varepsilon > 0$, and any instance of IIK with n items and T times, we can find a $(1 - \varepsilon)$ -approximation to a 1-in solution of highest profit in time $f(n, T, \varepsilon)$. Then we can find a $(1 - \varepsilon)$ -approximation to any instance of IIK with n items and T times in time $O(n) \cdot f(n, T, \varepsilon)$.*

Now, let \mathcal{I} be an instance of IIK with n items, let $\varepsilon > 0$. We say that \mathcal{I} is ε -well-behaved if it satisfies the following properties.

- ($\varepsilon 1$) For all $i \in [n]$, one has $p_i = (1 + \varepsilon)^{-j}$ for some $j \in [\lceil \log_{1+\varepsilon} \frac{T}{\varepsilon} \rceil]_0$, or $p_i \leq \frac{\varepsilon}{T}$.
- ($\varepsilon 2$) $b_t = b_{t-1}$ for all $t \in [T]$ such that $\lceil (1 + \varepsilon)^{j-1} \rceil < T - t + 1 < \lceil (1 + \varepsilon)^j \rceil$ for some $j \in [\lceil \log_{1+\varepsilon} T \rceil]$, where we set $b_0 = 0$.

See Figure 1 for an example. Note that condition ($\varepsilon 2$) implies that the capacity can change only during the set of times $\mathcal{T} := \{t \in [T] : t = T + 1 - \lceil (1 + \varepsilon)^j \rceil \text{ for some } j \in \mathbb{N}\}$, with $|\mathcal{T}| = O(\log_{1+\varepsilon} T)$. \mathcal{T} clearly gets sparser as t becomes smaller. Note also that times $t = 1, \dots, T - \lceil (1 + \varepsilon)^{\lceil \log_{1+\varepsilon} T \rceil} \rceil$ have capacity $b_t = 0$.

Next theorem implies that we can, wlog, assume that our instances are ε -well-behaved (and our solutions are 1-in).

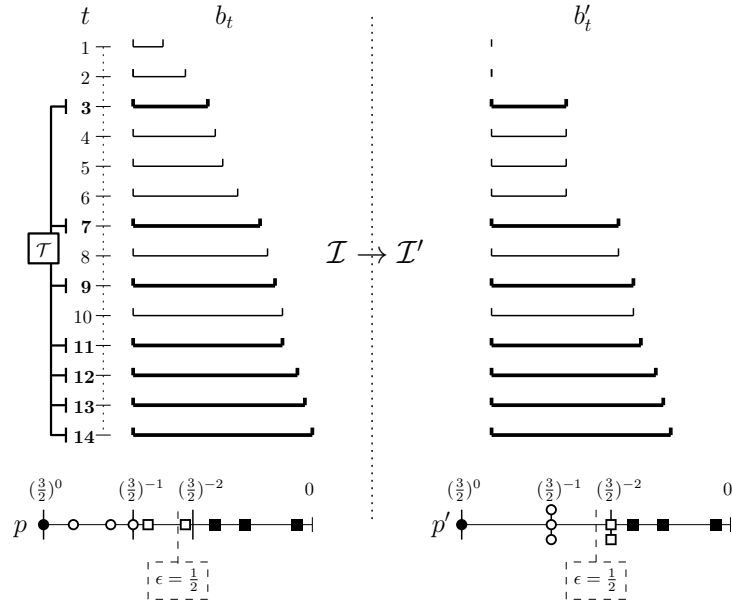


Figure 1: An example of obtaining an ε -well-behaved instance for $\varepsilon = \frac{1}{2}$ and $T = 14$.

Theorem 3. Suppose there exists a function $g : \mathbb{N} \times \mathbb{N} \times \mathbb{R}_{>0}$ such that, for each $n, T \in \mathbb{N}$, $\varepsilon > 0$, and any ε -well-behaved instance of IIK with n items and T times, we can find a $(1 - 3\varepsilon)$ -approximation to a 1-in solution of highest profit in time $g(n, T, \varepsilon)$. Then we can find a $(1 - 5\varepsilon)$ -approximation to any instance of IIK with n items and T times in time $O(T + n(n + g(n, T, \varepsilon)))$.

Proof. Fix an IIK instance \mathcal{I} . By Observation 1, it is enough to find a $(1 - 3\varepsilon)$ -approximation to its 1-in solution of highest profit in time $O(n + T) + g(n, T, \varepsilon)$. Consider instance \mathcal{I}' with n items having the same weights as in \mathcal{I} , T times, and the other parameters defined as follows:

- For $i \in [n]$, if $(1 + \varepsilon)^{-j} \leq p_i < (1 + \varepsilon)^{-j+1}$ for some $j \in [\lceil \log_{1+\varepsilon} \frac{T}{\varepsilon} \rceil]$, set $p'_i := (1 + \varepsilon)^{-j}$; otherwise, set $p'_i := p_i$. Note that we have $1 = p'_1 \geq p'_2 \geq \dots \geq p'_n$.
- For $t \in [T]$ and $\lceil (1 + \varepsilon)^{j-1} \rceil < T - t + 1 \leq \lceil (1 + \varepsilon)^j \rceil$ for some $j \in [\lceil \log_{1+\varepsilon} T \rceil]$, set $b'_t := b_{T - \lceil (1 + \varepsilon)^j \rceil + 1}$.
- For t such that $\lceil (1 + \varepsilon)^{\lceil \log_{1+\varepsilon} T \rceil} \rceil < T - t + 1 \leq T$ (i.e all the remaining t), set $b'_t := 0$.

One easily verifies that \mathcal{I}' is ε -well-behaved. Moreover, $b'_t \leq b_t$ for all $t \in [T]$ and $\frac{p_i}{1+\varepsilon} \leq p'_i \leq p_i$ for $i \in [n]$, so we deduce:

Claim 1. *Any solution \bar{x} feasible for \mathcal{I}' is also feasible for \mathcal{I} , and $p(\bar{x}) \geq p'(\bar{x})$.*

Claim 2. *Let x^* be a 1-in feasible solution of highest profit for \mathcal{I} . There exists a 1-in feasible solution x' for \mathcal{I}' such that $p'(x') \geq (1-\varepsilon)^2 p(x^*)$.*

Proof. Define $x' \in \{0,1\}^{Tn}$ as follows:

$$x'_t := \begin{cases} x_{T-\lceil(1+\varepsilon)^j+1}^* & \text{if } \lceil(1+\varepsilon)^{j-1}\rceil < T-t+1 \leq \lceil(1+\varepsilon)^j\rceil, j \in [\lceil\log_{1+\varepsilon} T\rceil], \\ 0 & \text{otherwise, i.e. for } t : \lceil(1+\varepsilon)^{\lceil\log_{1+\varepsilon} T\rceil}\rceil < T-t+1 \leq T. \end{cases}$$

In order to prove the claim we first show that x' is a feasible 1-in solution for \mathcal{I}' . Indeed, it is 1-in, since by construction $x'_{T,1} = x_{T,1}^* = 1$. It is feasible, since for t such that $\lceil(1+\varepsilon)^{j-1}\rceil < T-t+1 \leq \lceil(1+\varepsilon)^j\rceil$, $j \in \mathbb{N}, j \in [\lceil\log_{1+\varepsilon} T\rceil]$ we have

$$w^T x'_t = w^T x_{T-\lceil(1+\varepsilon)^j+1}^* \leq b_{T-\lceil(1+\varepsilon)^j+1} = b'_t,$$

while $w^T x'_t = 0 = b'_t$ otherwise.

Comparing $p'(x')$ and $p(x^*)$ gives

$$\begin{aligned} p'(x') &\geq \sum_{t \in [T]} \sum_{i \in [n]} p'_i x'_{t,i} &= \sum_{i \in [n]} (T - t_{i,\min}(x') + 1) p'_i \\ &\geq \sum_{i \in [n]} \frac{1}{1+\varepsilon} (T - t_{i,\min}(x^*) + 1) p'_i &\geq \sum_{i \in [n]} \frac{1}{(1+\varepsilon)^2} (T - t_{i,\min}(x^*) + 1) p_i \\ &= \left(\frac{1}{1+\varepsilon}\right)^2 p(x^*) &\geq (1-\varepsilon)^2 p(x^*), \end{aligned}$$

where $t_{i,\min}(v) := \min\{t \in [T] : v_{t,i} = 1\}$ for $v \in \{0,1\}^{Tn}$. (*End of the claim.*)

Let \hat{x} be a 1-in solution of highest profit for \mathcal{I}' and \bar{x} is a solution to \mathcal{I}' that is a $(1-\varepsilon)$ -approximation to \hat{x} . Claim 1 and Claim 2 imply that \bar{x} is feasible for \mathcal{I} and we deduce:

$$p(\bar{x}) \geq p'(\bar{x}) \geq (1-3\varepsilon)p'(\hat{x}) \geq (1-3\varepsilon)p'(x') \geq (1-3\varepsilon)(1-\varepsilon)^2 p(x^*) \geq (1-5\varepsilon)p(x^*).$$

In order to compute the running time, it is enough to bound the time required to produce \mathcal{I}' . Vector p' can be produced in time $O(n)$, while vector b' in time T . Moreover, the construction of the latter can be performed before fixing the highest profit object that belongs to the knapsack (see Observation 1). The thesis follows. \square

2.3 A disjunctive relaxation

Fix $\varepsilon > 0$. Because of Theorem 3, we can assume that the input instance \mathcal{I} is ε -well-behaved. We call all times from \mathcal{T} *significant*. Note that a solution over the latter times can be naturally extended to a global solution by setting $x_t = x_{t-1}$ for all non-significant times t . We denote significant times by $t_1 < t_2 < \dots < t_{|\mathcal{T}|}$.

In this section we describe an IP over feasible 1-in solutions of an ε -well-behaved instance of IIK. The feasible region of this IP is the union of different regions, each corresponding to a partial assignment of items to significant times. In Section 2.4 we give a strategy to round an optimal solution of the LP relaxation of the IP to a feasible integral solution with a $(1-3\varepsilon)$ -approximation guarantee. Together with Theorem 3 (taking $\varepsilon' = \frac{\varepsilon}{5}$), this implies Theorem 1.

In order to describe those partial assignments, we introduce some additional notation. We say that items having profit $(1+\varepsilon)^{-k}$ for $k \in [\lceil\log_{1+\varepsilon} \frac{T}{\varepsilon}\rceil]$, belong to *profit class* k . Hence bigger profit classes correspond to items with smaller profit. All other items are said to belong to the *small* profit class. Note that there are $O(\frac{1}{\varepsilon} \log \frac{T}{\varepsilon})$ profit classes (some of which could be empty). Our partial assignments will be induced by special sets of vertices of a related graph called *grid*.

Definition 4. Let $J \in \mathbb{Z}_{>0}, K \in \mathbb{Z}_{\geq 0}$, a grid of dimension $J \times (K + 1)$ is the graph $G_{J,K} = ([J] \times [K]_0, E)$, where

$$E := \{\{u, v\} : u, v \in [J] \times [K]_0, u = (j, k) \text{ and either } v = (j + 1, k) \text{ or } v = (j, k + 1)\}.$$

Definition 5. Given a grid $G_{J,K}$, we call $S = \{(j_1, k_1), (j_2, k_2), \dots, (j_{|S|}, k_{|S|})\} \subseteq V(G_{J,K})$ is a stairway if $j_h > j_{h+1}$ and $k_h < k_{h+1}$ for all $h \in [|S| - 1]$.

Lemma 6. There are at most 2^{K+1+J} distinct stairways in $G_{J,K}$.

Proof. The first coordinate of any entry of a stairway can be chosen among J values, the second coordinate from $K + 1$ values. By Definition 5, each stairway correspond to exactly one choice of sets $J_1 \subseteq [J]$ for the first coordinates and $K_1 \subseteq [K]_0$ for the second, with $|K_1| = |J_1|$. \square

Now consider the grid graph with $J := |\mathcal{T}| = O(\frac{1}{\varepsilon} \log T)$, $K = \lceil \log_{1+\varepsilon} \frac{T}{\varepsilon} \rceil$, and a stairway S with $k_1 = 0$. See Figure 2 for an example. This corresponds to a partial assignment that can be informally described as follows: if $(j_h, k_h) \in S$, then in the corresponding partial assignment no item belonging to profit classes $k_h \leq k < k_{h+1}$ is in the knapsack at any time $t < t_{j_h}$, while the first time an item from profit class k_h is inserted in the knapsack is time t_{j_h} (if $j_{|S|} > 1$ then items from the small profit class can only be placed in the knapsack at times $1, \dots, t_{j_{|S|}} - 1$). Moreover, for each $(\bar{j}, \bar{k}) \in S$, for an appropriately chosen family of profit classes k' following \bar{k} and significant times t' following $t_{\bar{j}}$, we will either specify exactly the number of items taken from k' at time t' , or impose that there are at least $\lceil \frac{1}{\varepsilon} \rceil$ of those items. Note that we can assume that the items taken within a profit class are those with minimum weight: this may exclude some feasible 1-in solutions, but it will always keep at least a feasible 1-in solution of maximum profit. No other constraint is imposed.

More formally, set $k_{|S|+1} = K + 1$, $C_\varepsilon = 2 \lceil \log_{1+\varepsilon} \frac{1}{\varepsilon} \rceil$. For each $h = 1, \dots, |S|$:

- i) Set $x_{t,i} = 0$ for all $t \in [t_{j_h} - 1]$ and each item i in a profit class $k \in [k_{h+1} - 1]$.
- ii) Define $\mathcal{T}_h := \{t_j \in \mathcal{T} : j_h \leq j \leq j_h^*\}$, where $j_h^* = \min\{j_h + C_\varepsilon, J\}$. For $k_h \leq k \leq \min\{k_h + C_\varepsilon, k_{h+1} - 1\}$, fix vectors $\rho_k \in \{0, 1, \dots, \lceil \frac{1}{\varepsilon} \rceil\}^{\mathcal{T}_h}$ such that $\rho_{t_j, k} \leq \rho_{t_{j+1}, k}$ for all $j : j_h \leq j < j_h^*$ and $\rho_{t_{j_h}, k_h} \geq 1$.

For each profit class $k \in [K]$ we assume that items $I_k = \{i_1^k, \dots, i_{|I_k|}^k\}$ in this class are ordered so that $w_{i_1^k} \leq w_{i_2^k} \leq \dots \leq w_{i_{|I_k|}^k}$. Based on our choice (S, ρ) we define the polytope:

$$P(S, \rho) = \{x \in \mathbb{R}^{Tn} : \quad w^T x_t \leq b_t \quad \forall t \in [T] \quad (2)$$

$$x_t \leq x_{t+1} \quad \forall t \in [T - 1] \quad (3)$$

$$0 \leq x_t \leq 1 \quad \forall t \in [T] \quad (4)$$

$$x_{t, i_1^k} = \dots = x_{t, i_{|I_k|}^k} = 0, \quad \forall h \in [|S|], \forall k < k_{h+1}, \forall t < t_{j_h} \quad (5)$$

$$x_{t_j, i_1^k} = \dots = x_{t_j, i_{\rho_{t_j, k}}^k} = 1, \quad \forall h \in [|S|], \forall t_j \in \mathcal{T}_h \quad (6)$$

$$x_{t_j, i_{(\rho_{t_j, k} + 1)}^k} = \dots = x_{t_j, i_{|I_k|}^k} = 0, \quad \forall h \in [|S|], \forall t_j \in \mathcal{T}_h : \rho_{t_j, k} < \frac{1}{\varepsilon}. \quad (7)$$

Note that some choices of S, ρ may lead to empty polytopes. Fix S, ρ , an item i and some time t . If, for some $t' \leq t$, $x_{t', i} = 1$ explicitly appears in the definition of $P(S, \rho)$ above, then we say that i is t -included. Conversely, if $x_{\bar{t}, i} = 0$ explicitly appears for some $\bar{t} \geq t$, then we say that i is t -excluded.

Theorem 7. Any optimum solution of $\max\{\sum_{t \in [T]} p_t^T x_t : x \in (\cup_{S, \rho} P(S, \rho)) \cap \{0, 1\}^{Tn}\}$ is a 1-in solution of maximum profit for \mathcal{I} . Moreover, the the number of constraints of the associated LP relaxation is at most $nT^{f(\varepsilon)}$ for some function $f : \mathbb{R}_{>0} \rightarrow \mathbb{R}_{>0}$ depending on ε only.

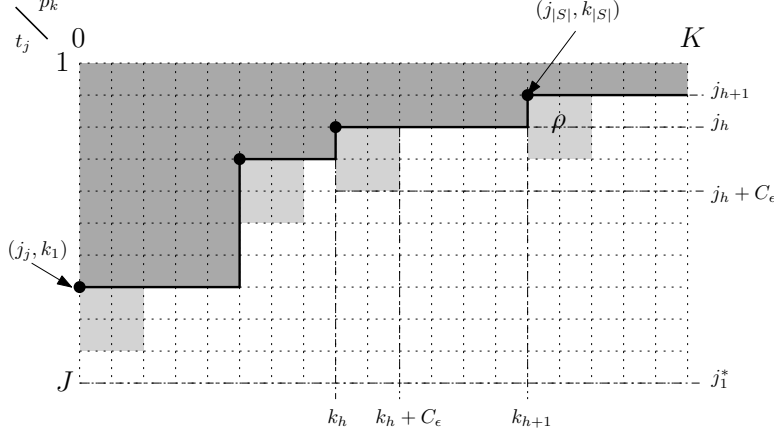


Figure 2: An example of a stairway S , given by thick black dots. Entries (j, k) lying in the light grey are those for which a value ρ is specified. No item corresponding to the entries in the dark grey is taken, except on the boundary in bold.

Proof. The first part of the statement follows from the previous discussion. The second from the fact that the possible choices of (S, ρ) are

$$\begin{aligned}
 (\# \text{ stairways}) &\cdot (\# \text{ possible values in each entry of } \rho) \quad (\max \# \text{ entries of a vector } \rho) \\
 &= \\
 2^{O(\frac{1}{\varepsilon} \log \frac{T}{\varepsilon})} &\cdot O(\frac{1}{\varepsilon}) \quad O(\frac{1}{\varepsilon} \log \frac{T}{\varepsilon})(C_\varepsilon)^2 \\
 &= \\
 (\frac{T}{\varepsilon})^{O(\frac{1}{\varepsilon})} &\cdot (\frac{T}{\varepsilon})^{O((\frac{1}{\varepsilon})^4)},
 \end{aligned}$$

and each of them has $g(\varepsilon)O(Tn)$ constraints, where g depends on ε only. \square

2.4 Rounding

By convexity, there is a choice of S and ρ as in the previous section such that any optimum solution of

$$\max \left\{ \sum_{t \in [T]} p^T x_t : x \in P(S, \rho) \right\} \quad (8)$$

is also an optimum solution to $\max \{ \sum_{t \in [T]} p^T x_t : x \in \text{conv}(\cup_{S, \rho} P(S, \rho)) \}$. Hence, we can focus on rounding an optimum solution x^* of (8). For $h \in [|S|]$, $\mathcal{O}_h := \{i : i \text{ is in profit class } k \text{ for some } k \in \mathcal{K}_h\}$ with $\mathcal{K}_h := \{k_h \leq k < k_{h+1}\}$, and $\mathcal{O}_\infty := [n] \setminus \cup_{h=1}^{|S|} \mathcal{O}_h$. Hence, \mathcal{O}_∞ is the set of items from the small profit class. Let $o_1, o_2, \dots, o_{|\mathcal{O}_h|}$ be items from \mathcal{O}_h sorted by decreasing value of their profit / weight ratio. Moreover, let \mathcal{I}_h^t (resp. \mathcal{E}_h^t) be the set of items from \mathcal{O}_h that are t -included (resp. t -excluded) and, for $t \in [T]$, let $W_t := \sum_{i \in \mathcal{O}_h} w_i x_{t,i}^*$. Algorithm 1 produces, for each $h \in [|S|] \cup \{\infty\}$, a value $\bar{x}_{t,i}$ for $t \in [T]$ and $i \in \mathcal{O}_h$. Respecting the choices of S and ρ , at each time t Algorithm 1 greedily adds objects into the knapsack, until the total weight is at most W_t . The juxtaposition \bar{x} of those vectors is the claimed approximated integer solution. As in MAX-K we aim at obtaining a rounded solution which differs from x^* by profit of at most one item (at each time). However, the structure of x^* is much more subtle than the optimal fractional solution of MAX-K.

Theorem 8. *Let x^* be an optimum solution to (8). Apply Algorithm 1 for each $h \in [|S|] \cup \{\infty\}$, as to produce, in time $O(T + n)$, an integer vector \bar{x} . Then $x \in P(S, \rho)$ and $\sum_{t \in [T]} p^T \bar{x}_t \geq (1 - 3\varepsilon) \sum_{t \in [T]} p^T x_t^*$.*

Algorithm 1

- 1: For all $i \in \mathcal{O}_h$, set $\bar{x}_{0,i} = 0$.
 - 2: For $t = 1, \dots, T$:
 - (a) Set $\bar{x}_t = \bar{x}_{t-1}$.
 - (b) Set $\bar{x}_{t,i} = 1$ for all $i \in \mathcal{I}_h^t$.
 - (c) Select the smallest $p \in \mathbb{N}$ such that $o_p \notin \mathcal{E}_h^t$ and $\bar{x}_{t,o_p} = 0$.
 - (d) If $w_{o_p} \leq W_t - \sum_{i \in \mathcal{O}_h} w_i \bar{x}_{t,i}$, set $\bar{x}_{t,o_p} = 1$ and go to (c), else stop.
-

Theorem 8 will be proved in a series of intermediate steps. Until differently specified, we suppose that $h \in [|S|] \cup \{\infty\}$ is fixed.

Claim 3. *Let $t \in [T - 1]$. Then:*

- (i) $\mathcal{I}_h^t \subseteq \mathcal{I}_h^{t+1}$ and $\mathcal{E}_h^t \supseteq \mathcal{E}_h^{t+1}$.
- (ii) $\mathcal{I}_h^{t+1} \setminus \mathcal{I}_h^t \subseteq \mathcal{E}_h^t$.

Proof. (i) Immediately from the definition.

- (ii) If $\mathcal{I}_h^{t+1} \setminus \mathcal{I}_h^t \neq \emptyset$, we deduce $t + 1 = t_j$ for some $j_h \leq j \leq j_{h^*}$. By construction, the items $\mathcal{I}_h^{t+1} \setminus \mathcal{I}_h^t$ can only be in buckets $k : k_h \leq k < k_{h+1}$ where $\rho_{t,k} < \lceil \frac{1}{\varepsilon} \rceil$. Hence, all items from $\mathcal{I}_h^{t+1} \setminus \mathcal{I}_h^t$ are t -excluded. □

Recall that, for $t \in [T]$, let $W_t := \sum_{i \in \mathcal{O}_h} w_i x_{t,i}^*$. Note that $W_1 = W_2 = \dots = W_{t_{j_h}-1} = 0$ and $W_t \leq W_{t+1}$ for all $t \in [T - 1]$. Algorithm 2 provides a constructive way to produce the restriction to \mathcal{O}_h of an optimum solution to (8).

Algorithm 2

- 1: For all $i \in \mathcal{O}_h$, set $x'_{0,i} = 0$.
 - 2: For $t = 1, \dots, T$:
 - (a) Set $x'_t = x'_{t-1}$.
 - (b) Set $x'_{t,i} = 1$ for all $i \in \mathcal{I}_h^t$.
 - (c) While $W_t - \sum_{i \in \mathcal{O}_h} w_i x'_{t,i} > 0$:
 - (i) Select the smallest $p \in \mathbb{N}$ such that $o_p \notin \mathcal{E}_h^t$ and $x'_{t,o_p} < 1$.
 - (ii) Set $x'_{t,o_p} = x'_{t,o_p} + \min\{1 - x'_{t,o_p}, \frac{W_t - \sum_{i \in \mathcal{O}_h} w_i x'_{t,i}}{w_{o_p}}\}$.
-

The proof of the following claim easily follows by construction.

Claim 4. (i) *For $t \in [t_{j_h} - 1]$ and $i \in \mathcal{O}_h$, one has $x_{t,i}^* = x'_{t,i} = 0$.*

(ii) *For $t \in [T - 1]$ and $i \in \mathcal{O}_h$, one has $x'_{t+1,i} \geq x'_{t,i} \geq 0$.*

(iii) For $t \in [T]$, one has: $x_{t,i}^* = x'_{t,i} = 1$ for $i \in \mathcal{I}_h^t$ and $x_{t,i}^* = x'_{t,i} = 0$ for $i \in \mathcal{E}_h^t$.

Claim 5. Let x' be the solution produced by Algorithm 2. Then for each $t \in [T]$, $\sum_{i \in \mathcal{O}_h} w_i x'_{t,i} = \sum_{i \in \mathcal{O}_h} w_i x_{t,i}^*$ and $\sum_{i \in \mathcal{O}_h} p_i x'_{t,i} = \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^*$.

Proof. We first prove the statement on the weights by induction on t , the basic step being trivial. Suppose it is true up to time $t-1$. The total weight of solution x'_t after step (b) is

$$\begin{aligned} \sum_{i \in \mathcal{O}_h} w_i x'_{t-1,i} + \sum_{i \in \mathcal{I}_h^t \setminus \mathcal{I}_h^{t-1}} w_i (1 - x'_{t-1,i}) &= W_{t-1} + \sum_{i \in \mathcal{I}_h^t \setminus \mathcal{I}_h^{t-1}} w_i (1 - x_{t-1,i}^*) \\ &= W_{t-1} + \sum_{i \in \mathcal{I}_h^t \setminus \mathcal{I}_h^{t-1}} w_i \stackrel{(*)}{\leq} W_t, \end{aligned}$$

where the equations follow by induction, Claim 4.(iii), and Claim 3.(ii), and $(*)$ follows by observing $\sum_{i \in \mathcal{O}_h} w_i x_{t,i}^* - w_i x_{t-1,i}^* \geq \sum_{i \in \mathcal{I}_h^t \setminus \mathcal{I}_h^{t-1}} w_i$. x'_t is afterwards increased until its total weight is at most W_t . Last, observe that W_t is always achieved, since it is achieved by x_t^* . This concludes the proof of the first statement.

We now move to the statement on profits. Note that it immediately follows from the optimality of x^* and the first part of the claim if we show that x' is the solution maximizing $p^T x_t$ for all $t \in [T]$, among all $x \in P(S, \rho)$ that satisfy $\sum_{i \in \mathcal{O}_h} w_i x_{t,i} = W_t$ for all $t \in [T]$. So let us prove the latter. Suppose by contradiction this is not the case, and let \tilde{x} be one such solution such that $p^T \tilde{x}_t > p^T x'_t$ for some $t \in [T]$. Among all such \tilde{x} , take one that is lexicographically maximal, where entries are ordered $(1, o_1), (1, o_2), \dots, (1, o_{|\mathcal{O}_h|}), (2, o_1), \dots, (T, o_{|\mathcal{O}_h|})$. Then there exists $\tau \in [T]$, $\ell \in [|\mathcal{O}_h|]$ such that $\tilde{x}_{\tau, o_\ell} > x'_{\tau, o_\ell}$. Pick τ minimum such that this happens, and ℓ minimum for this τ . Using that $x'_{\tau, i} = \tilde{x}_{\tau, i}$ for $i \in \mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau$ since $x', \tilde{x} \in P(S, \rho)$ and recalling $\sum_{i \in \mathcal{O}_h} w_i x'_{\tau, i} = \sum_{i \in \mathcal{O}_h} w_i \tilde{x}_{\tau, i} = W_\tau$ one obtains

$$\sum_{i \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau)} w_i x'_{\tau, i} = \sum_{i \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau)} w_i \tilde{x}_{\tau, i} \quad (9)$$

It must be that $x'_{\tau, o_\ell} < 1$, since $x'_{\tau, o_\ell} < \tilde{x}_{\tau, o_\ell} \leq 1$, so step (c) of Algorithm 2 in iteration τ did not change any item $o_{\hat{\ell}} : \hat{\ell} > \ell$, i.e. $x'_{\tau, o_{\hat{\ell}}} = x'_{\tau-1, o_{\hat{\ell}}}$ for each $\hat{\ell} > \ell$. Additionally, $\ell \notin \mathcal{I}_h^\tau$ because $x'_{\tau, o_\ell} < 1$, and $\ell \notin \mathcal{E}_h^\tau$ since otherwise $x'_{\tau, o_\ell} = \tilde{x}_{\tau, o_\ell} = 0$. Hence, $\ell \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau)$. We rewrite (9) as follows:

$$\sum_{\substack{o_{\hat{\ell}} \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau): \\ \hat{\ell} \leq \ell}} w_{o_{\hat{\ell}}} x'_{\tau, o_{\hat{\ell}}} = \sum_{\substack{o_{\hat{\ell}} \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau): \\ \hat{\ell} \leq \ell}} w_{o_{\hat{\ell}}} \tilde{x}_{\tau, o_{\hat{\ell}}} + \sum_{\substack{o_{\hat{\ell}} \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau): \\ \hat{\ell} > \ell}} w_{o_{\hat{\ell}}} (\tilde{x}_{\tau, o_{\hat{\ell}}} - \underbrace{x'_{\tau, o_{\hat{\ell}}}}_{=x'_{\tau-1, o_{\hat{\ell}}}})$$

By minimality of τ one has $\tilde{x}_{\tau-1} \leq x'_{\tau-1}$, so $\sum_{i \in \mathcal{O}_h} w_i \tilde{x}_{\tau-1, i} = W_{\tau-1} = \sum_{i \in \mathcal{O}_h} w_i x'_{\tau-1, i}$ implies $\tilde{x}_{\tau-1} = x'_{\tau-1}$ and thus

$$\sum_{\substack{o_{\hat{\ell}} \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau): \\ \hat{\ell} \leq \ell}} w_{o_{\hat{\ell}}} x'_{\tau, o_{\hat{\ell}}} = \sum_{\substack{o_{\hat{\ell}} \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau): \\ \hat{\ell} \leq \ell}} w_{o_{\hat{\ell}}} \tilde{x}_{\tau, o_{\hat{\ell}}} + \overbrace{\sum_{\substack{o_{\hat{\ell}} \in \mathcal{O}_h \setminus (\mathcal{I}_h^\tau \cup \mathcal{E}_h^\tau): \\ \hat{\ell} > \ell}} w_{o_{\hat{\ell}}} (\tilde{x}_{\tau, o_{\hat{\ell}}} - \tilde{x}_{\tau-1, o_{\hat{\ell}}})}^{\geq 0} \quad (10)$$

Note that the items in \mathcal{O}_h are ordered according to monotonically decreasing profit/weight ratio. By minimality of ℓ subject to τ we have that $x'_{\tau, o_{\hat{\ell}}} \geq \tilde{x}_{\tau, o_{\hat{\ell}}}$ for $\hat{\ell} < \ell$. Thus combining $x'_{\tau, o_\ell} < \tilde{x}_{\tau, o_\ell}$ with (10) gives that there exists $\beta < \ell$ such that $x'_{\tau, o_\beta} > \tilde{x}_{\tau, o_\beta}$. Then for all $\bar{\tau} \geq \tau$, one can perturb \tilde{x} by increasing $\tilde{x}_{\bar{\tau}, o_\beta}$ and decreasing $\tilde{x}_{\bar{\tau}, o_\ell}$ while keeping $\tilde{x} \in P(S, \rho)$ and $\sum_{i \in \mathcal{O}_h} w_{\bar{\tau}, i} \tilde{x}_{\bar{\tau}, i} = W_{\bar{\tau}}$, without decreasing $p^T \tilde{x}_{\bar{\tau}}$. This contradicts the choice of \tilde{x} being lexicographically maximal. \square

Because of Claim 5, we suppose wlog $x_{t,i}^* = x'_{t,i}$ for $t \in [T]$ and $i \in \mathcal{O}_h$. Note that Algorithm 1 can be seen as a “discrete version” of Algorithm 2.

Claim 6. (i) For $t \in [T]$ and $i \in \mathcal{O}_h$, one has $\bar{x}_{t,i} \in \{0, 1\}$.

(ii) For $t \in [T-1]$ and $i \in \mathcal{O}_h$, one has $\bar{x}_{t+1,i} \geq \bar{x}_{t,i}$.

(iii) For $t \in [T]$, one has: $\bar{x}_{t,i} = 1$ for $i \in \mathcal{I}_h^t$ and $\bar{x}_{t,i} = 0$ for $i \in \mathcal{E}_h^t$.

(iv) For $t \in [T]$, one has $\sum_{i \in \mathcal{O}_h} w_i \bar{x}_{t,i} \leq W_t$.

(v) For $t \in [T]$, one has $\sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} \geq \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* - \max_{i \in \mathcal{O}_h \setminus (\mathcal{I}_h^t \cup \mathcal{E}_h^t)} p_i$.

Proof.

(i)-(iii) are immediate, and (iv) proceeds similarly to the proof of the statement on weights from Claim 5, so we omit it.

(v) Fix $t \in [T]$, and let first $x_{t,o_\ell}^* = 1$. We claim that $\bar{x}_{t,o_\ell} = 1$. This follows from Part (iii) if $o_\ell \in \mathcal{I}_h^t$. So suppose $o_\ell \in \mathcal{O}_h \setminus (\mathcal{I}_h^t \cup \mathcal{E}_h^t)$. Consider the minimum τ such that $x_{\tau,o_\ell}^* = 1$. Then all items o_i with $i < \ell$ either satisfy $x_{\tau,o_i}^* = 1$, or $o_i \in \mathcal{E}_h^\tau$, hence $\sum_{i \leq \ell: o_i \notin \mathcal{E}_h^\tau} w_{o_i} \leq W_\tau$. Hence, by construction, Algorithm 1 sets $\bar{x}_{\tau,\ell} = 1$. The claim then follows from part (ii).

Now let F be the set of indices j such that o_j is not t -excluded and $x_{t,o_j}^* \neq 1$, and let $r \in F$ be such that $\sum_{j \in F} w_{o_j} x_{t,o_j}^* < \sum_{j \in F: j \leq r} w_{o_j}$ and minimum with this property. Then

$$\begin{aligned} W_t &= \sum_{i \in \mathcal{O}_h} w_i x_{t,i}^* = \sum_{i: x_{t,i}^* = 1} w_i + \sum_{j \in F} w_{o_j} x_{t,o_j}^* \\ &\in [\sum_{i: x_{t,i}^* = 1} w_i + \sum_{j \in F: j \leq r-1} w_{o_j}, \sum_{i: x_{t,i}^* = 1} w_i + \sum_{j \in F: j \leq r} w_{o_j}] \end{aligned}$$

where we used the definition of r . Then, because of what we proved above and by construction, Algorithm 1 sets $\bar{x}_{t,o_j} = 1$ for $j : x_{o_j}^* = 1$, and $j \in F$ such that $j \leq r-1$. We have therefore

$$\begin{aligned} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* - \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} &\leq \sum_{j \in F} p_{o_j} x_{t,o_j}^* - \sum_{j \in F: j \leq r-1} p_{o_j} \\ &= \sum_{j \in F: j \geq r} p_{o_j} x_{t,o_j}^* - \sum_{j \in F: j \leq r-1} p_{o_j} (1 - x_{t,o_j}^*) \\ &\leq \frac{p_{o_r}}{w_{o_r}} (\sum_{j \in F: j \geq r} w_{o_j} x_{t,o_j}^* - \sum_{j \in F: j \leq r-1} w_{o_j} (1 - x_{t,o_j}^*)) \\ &= \frac{p_{o_r}}{w_{o_r}} (\sum_{j \in F} w_{o_j} x_{o_j}^* - \sum_{j \in F: j \leq r-1} w_{o_j}) < \frac{p_{o_r}}{w_{o_r}} w_{o_r} = p_{o_r}, \end{aligned}$$

concluding the proof. \square

We now consider the vector \bar{x} obtained by juxtaposition of the vectors produced by Algorithm 1 for each $h \in [S] \cup \{\infty\}$.

Claim 7. Let $h \in [S]$ and $t \in [t_{j_h}, \dots, t_{j_h}^*]$. Then $\sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} \geq (1 - \varepsilon) \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^*$.

Proof. Fix t as in the statement of the claim. We can assume $\sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} < \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^*$, else there is nothing to prove. We apply Claim 6.(v) and distinguish two cases according to the item achieving $\max_{i \in \mathcal{O}_h \setminus (\mathcal{I}_h^t \cup \mathcal{E}_h^t)} p_i$. Suppose first it belongs to a profit class $k : k_h \leq k < \min\{k_h + C_\varepsilon + 1, k_{h+1}\}$. Since it is not t -included or t -excluded, we must have that $\rho_{k,t} = \frac{1}{\varepsilon}$. Hence,

$$\max_{i \in \mathcal{O}_h \setminus (\mathcal{I}_h^t \cup \mathcal{E}_h^t)} p_i \leq \varepsilon \sum_{i \in \mathcal{I}_h^t} p_i \bar{x}_{t,i} \leq \varepsilon \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i}.$$

On the other hand, if it belongs to a profit class $k : k_h + C_\varepsilon < k < k_{h+1}$, then we have again

$$\max_{i \in \mathcal{O}_h \setminus (\mathcal{I}_h^t \cup \mathcal{E}_h^t)} p_i \leq (1 + \varepsilon)^{-C_\varepsilon} (1 + \varepsilon)^{-k_h} \leq \varepsilon (1 + \varepsilon)^{-k_h} \leq \varepsilon \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i},$$

where last inequality follows from the fact that $\rho_{k_h,t} \geq 1$ by construction. In both cases, the thesis follows by the assumption $\sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} < \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^*$. \square

Claim 8. $\sum_{t \in [T]} \sum_{i \in \mathcal{O}_\infty} p_i \bar{x}_{t,i} \geq \sum_{t \in [T]} \sum_{i \in \mathcal{O}_\infty} p_i x_{t,i}^* - \varepsilon \sum_{t \in [T]} p^T x_t^*.$

Proof. $\max_{i \in \mathcal{O}_\infty} p_i \leq \frac{\varepsilon}{T} = \frac{\varepsilon}{T} p_1 \leq \frac{\varepsilon}{T} \sum_{t \in [T]} p^T x_t^*$, where the first inequality follows from the definition of small profit class and the last from the fact that, by definition of ε -well-behaved, $x_{T,1}^* = 1$. The statement then follows from Claim 6.(v). \square

Claim 9. Let $h \in [|S|]$. Then $\sum_{t > t_{j_h}^*} \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} \geq \sum_{t > t_{j_h}^*} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* - \varepsilon \sum_{t \in [T]} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^*.$

Proof. Let ℓ be the highest profit item from \mathcal{O}_h , and recall that, by construction, $\ell \in \mathcal{I}_h^{t_{j_h}}$. Hence, $\sum_{t \in [T]} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* \geq (T - t_{j_h} + 1) p_\ell = (T - t_{j_h}^* + 1) \frac{[(1+\varepsilon)^{J-j_h}]}{[(1+\varepsilon)^{J-j_h} - \varepsilon]} p_\ell \geq (T - t_{j_h}^* + 1) \frac{(1+\varepsilon)^{J-j_h}}{(1+\varepsilon)^{J-j_h} \cdot \varepsilon \cdot (1-(1-\varepsilon)) + 1} p_\ell \geq (T - t_{j_h}^* + 1) \frac{(1+\varepsilon)^{J-j_h}}{(1+\varepsilon)^{J-j_h} \cdot \varepsilon \cdot \frac{1}{\varepsilon^2} \varepsilon (1-\varepsilon) + 1} p_\ell \geq \frac{T - t_{j_h}^*}{\varepsilon} \max_{i \in \mathcal{O}_h} p_i$ with $\frac{1}{\varepsilon}(1-\varepsilon) \geq 1$ for $\varepsilon \leq 1/2$. The statement then follows from Claim 6.(v). \square

We now have all the ingredients to prove Theorem 8.

Proof of Theorem 8. By Claim 6.(i)-(iii), we have that $\bar{x} \in P(S, \rho) \cap \{0, 1\}^{Tn}$. Moreover:

$$\begin{aligned}
\sum_{t \in [T]} p^T \bar{x}_t &= \sum_{t \in [T]} \sum_{h \in [|S|] \cup \{\infty\}} \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} \\
&= \sum_{h \in [|S|]} \left(\sum_{t \in [t_{j_h}, \dots, t_{j_h}^*]} \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} + \sum_{t > t_{j_h}^*} \sum_{i \in \mathcal{O}_h} p_i \bar{x}_{t,i} \right) + \\
&\quad + \sum_{t \in [T]} \sum_{i \in \mathcal{O}_\infty} p_i \bar{x}_{t,i} \\
(\text{Using Claim 7, 8, 9}) &\geq \sum_{h \in [|S|]} \left((1 - \varepsilon) \sum_{t \in [t_{j_h}, \dots, t_{j_h}^*]} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* + \sum_{t > t_{j_h}^*} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* + \right. \\
&\quad \left. - \varepsilon \sum_{t \in [T]} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* \right) + \sum_{t \in [T]} \sum_{i \in \mathcal{O}_\infty} p_i x_{t,i}^* - \varepsilon \sum_{t \in [T]} p^T x_t^* \\
&= \sum_{h \in [|S|] \cup \{\infty\}} \sum_{t \in [T]} \sum_{i \in \mathcal{O}_h} p_i x_{t,i}^* + \\
&\quad - \sum_{h \in [|S|]} \sum_{i \in \mathcal{O}_h} \left(\varepsilon \sum_{t \in [t_{j_h}, \dots, t_{j_h}^*]} p_i x_{t,i}^* + \varepsilon \sum_{t \in [T]} p_i x_{t,i}^* \right) - \varepsilon \sum_{t \in [T]} p^T x_t^* \\
&\geq \sum_{t \in [T]} p^T x_t^* - 2\varepsilon \sum_{h \in [|S|]} \sum_{i \in \mathcal{O}_h} \sum_{t \in [T]} p_i x_{t,i}^* - \varepsilon \sum_{t \in [T]} p^T x_t^* \\
&\geq \sum_{t \in [T]} p^T x_t^* - 3\varepsilon \sum_{t \in [T]} p^T x_t^*,
\end{aligned}$$

as required.

Theorem 1 now easily follows from Theorems 3, 7, and 8.

Proof of Theorem 1. Since we will need items to be sorted by profit / weight ratio, we can do this once and for all before any guessing is performed. Classical algorithms implement this in $O(n \log n)$. By Theorem 3, we know we can assume that the input instance is ε -well-behaved, and it is enough to find a solution of profit at least $(1 - 3\varepsilon)$ the profit of a 1-in solution of maximum profit - By Theorem 8, this is exactly the vector \bar{x} . In order to produce \bar{x} , as we already sorted items by profit / weight ratio, we only need to solve the LPs associated with each choice of S and ρ , and then run Algorithm 1. The number of choices of S and ρ are $T^{f(\varepsilon)}$, and each LP has $g(\varepsilon)O(nT)$ constraints, for appropriate functions f and g (see the proof of Theorem 7). Algorithm 2 runs in time $O(\frac{T}{\varepsilon} \log \frac{T}{\varepsilon} + n)$. The overall running time is:

$$O(n \log n + n(n + T + T^{f(\varepsilon)}(f_{LP}(g(\varepsilon)O(nT)) + \frac{T}{\varepsilon} \log \frac{T}{\varepsilon}))) = O(nT^{h(\varepsilon)} f_{LP}(n)),$$

where $f_{LP}(m)$ is the time required to solve an LP with $O(m)$ variables and constraints, and $h : \mathbb{R} \rightarrow \mathbb{N}_{\geq 1}$ is an appropriate function.

2.5 Generalizations

Following Theorem 1, one could ask for a PTAS for the general incremental knapsack (IK) problem. This is the modification of IIK (introduced in [8]) where the objective function is $p_\Delta(x) := \sum_{t \in [T]} \Delta_t \cdot p^T x_t$, where $\Delta_t \in \mathbb{Z}_{>0}$ for $t \in [T]$ can be seen as time-dependent discounts. We show here some partial results.

Corollary 9. *There exists a PTAS-preserving reduction from IK to IIK, assuming $\Delta_t \leq \Delta_{t+1}$ for $t \in [T-1]$. Hence, the hypothesis above, IK has a PTAS.*

We start by proving an auxiliary corollary.

Corollary 10. *There exists a strict approximation-preserving reduction from IK to IIK, assuming that the maximum discount $\Delta_{\max} := \|\Delta\|_\infty$ is bounded by a polynomial $g(T, n, \log \|p\|_\infty, \log \|w\|_\infty)$. In particular, under the hypothesis above, IK has a PTAS.*

Proof. Let $\mathcal{I} := (n, p, w, T, b, \Delta)$ be an instance of IK with $\Delta_{\max} \leq g(T, n, \log \|p\|_\infty, \log \|w\|_\infty)$. The corresponding instance $\mathcal{I}' := (n, p, w, T', b')$ of IIK is obtained by setting $T' := \sum_{t \in [T]} \Delta_t$ and $b'_{t'} := b_t$ for $t' \in [T']$ if $\delta_t + 1 \leq t' \leq \delta_t + \Delta_t$, where $\delta_t := \sum_{\bar{t} < t} \Delta_{\bar{t}}$ for $t \in [T]$. We have that $T' \leq T \cdot g(T, n, \log \|p\|_\infty, \log \|w\|_\infty)$ so the size of \mathcal{I}' is polynomial in the size of \mathcal{I} .

Given an optimal solution $x^* \in \{0, 1\}^{Tn}$ to \mathcal{I} , and $x' \in \{0, 1\}^{T'n}$ such that $x'_{t'} = x_t$ for all $t \in [T]$ and $\delta_t + 1 \leq t' \leq \delta_t + \Delta_t$, one has that x' is feasible in \mathcal{I}' so

$$\text{OPT}(\mathcal{I}) = p_\Delta(x^*) = \sum_{t \in [T]} \Delta_t \cdot p^T x_t^* = \sum_{t' \in [T']} p^T x'_{t'} \leq \text{OPT}(\mathcal{I}').$$

Let \hat{x} be a α -approximated solution to \mathcal{I}' . Define $\bar{x} \in \{0, 1\}^{Tn}$ as $\bar{x}_t = \hat{x}_{\delta_t + \Delta_t}$ for $t \in [T]$. Then clearly $\bar{x}_t \leq \bar{x}_{t+1}$ for $t \in [T-1]$. Moreover, $w^T \bar{x}_t = w^T \hat{x}_{\delta_t + \Delta_t} \leq b'_{\delta_t + \Delta_t} = b_t$ for each $t \in [T]$. Hence \bar{x} is a feasible solution for \mathcal{I} and $p_\Delta(\bar{x}) = \sum_{t \in [T]} \Delta_t \cdot p^T \bar{x}_t \geq \sum_{\bar{t} \in [T']} p^T \hat{x}_{\bar{t}}$. Finally, one obtains:

$$\frac{p_\Delta(\bar{x})}{\text{OPT}(\mathcal{I})} \geq \frac{\sum_{\bar{t} \in [T']} p^T \hat{x}_{\bar{t}}}{\text{OPT}(\mathcal{I}')} \geq \alpha.$$

□

Proof of Corollary 9. Given an instance \mathcal{I} of IK with monotonically increasing discounts, and letting $p_{\max} := \|p\|_\infty$, we have that the optimal solution of \mathcal{I} is at least $\Delta_{\max} \cdot p_{\max}$ since $w_i \leq b_T$, $\forall i \in [n]$, otherwise an element i can be discarded from the consideration. Reduce \mathcal{I} to an instance \mathcal{I}' by setting $C = \frac{\varepsilon \Delta_{\max}}{Tn}$ and $\Delta'_t = \lfloor \frac{\Delta_t}{C} \rfloor$. We get that $\Delta'_{\max} \leq Tn/\varepsilon$ thus satisfying the assumption of Corollary 10 for each fixed $\varepsilon > 0$. Let x^* be an optimal solution to \mathcal{I} and \bar{x} a $(1 - \varepsilon)$ -approximated solution to \mathcal{I}' , one has:

$$\begin{aligned} p_\Delta(\bar{x}) &\geq C \cdot p'_\Delta(\bar{x}) \\ &\geq C \cdot (1 - \varepsilon) p'_\Delta(x^*) \\ &\geq (1 - \varepsilon) (p_\Delta(x^*) - C \sum_t p^T x_t^*) \\ &\geq (1 - \varepsilon) (p_\Delta(x^*) - \varepsilon \Delta_{\max} \cdot p_{\max}) \geq (1 - 2\varepsilon) p_\Delta(x^*). \end{aligned}$$

The proof of Corollary 9 only uses the fact that an item of the maximum profit is feasible at a time with the highest discount. Thus its implications are broader. Of independent interest is the fact that there is a PTAS for the modified version of IIK when each item can be taken multiple times. Unlike Corollary 9, this is not based on a reduction between problems, but on a modification of our algorithm.

Corollary 11. *There is a PTAS for the following modification of IIK: in (1), replace $x_t \in \{0, 1\}^n$ with: $x_t \in \mathbb{Z}_{>0}^n$ for $t \in [T]$; and $0 \leq x_t \leq d$ for $t \in [T]$, where we let $d \in (\mathbb{Z}_{>0} \cup \{\infty\})^n$ be part of the input.*

Proof. We detail the changes to be implemented to the algorithm and omit the analysis, since it closes follows that for IIK.

Modify the definition of $P(S, \rho)$ as follows. Fix $h \in [|S|]$, $k \in [K]$ such that $k_h \leq k \leq \min\{k_h + C_\varepsilon, k_{h+1} - 1\}$, and $t := t_j, j : j_h \leq j \leq j_h^*$. As before, items in the k -th bucket are ordered monotonically increasing according to their weight as $I_k = \{i_1, \dots, i_{|I_k|}\}$. In order to encounter item multiplicities we define $r := \max\{\bar{r} : \sum_{l=1}^{\bar{r}} d_{i_l} < \rho_{t,k}\}$. We change the following constraints of $P(S, \rho)$:

$$(4') \quad 0 \leq x_t \leq d.$$

$$(5') \quad x_{t,i_1} = d_{i_1}, \dots, x_{t,i_r} = d_{i_r}. \text{ Let } \bar{d} := \rho_{t,k} - \sum_{l=1}^r d_{i_l}. \text{ If } \rho_{t,k} < \frac{1}{\varepsilon} \text{ we set } x_{t,j,i_{r+1}} = \bar{d}, \text{ and } x_{t,j,i_{r+1}} \geq \bar{d} \text{ otherwise.}$$

$$(6') \quad \text{If } \rho_{t,k} < \frac{1}{\varepsilon} \text{ we set } x_{t,i_{r+2}} = 0, \dots, x_{t,i_{|I_k|}} = 0.$$

For fixed S, ρ , call all items i such that $x_{t,i} = c$ appears in (5') or in (6') (t, c) -fixed (note that items that are $(t, 0)$ -fixed correspond to items that were called t -excluded in IIK). Items that are (t, c) -fixed for some c are called t -fixed. The following modification of Algorithm 2 gives the structure of the optimal fractional solution to a fixed $P(S, \rho)$:

(b') For $i \in [n]$, if i is (t, c) -fixed, set $x'_{t,i} = c$.

(c') While $W_t - \sum_{i \in \mathcal{O}_h} w_i x'_{t,i} > 0$:

(i) Select the smallest $p \in \mathbb{N}$ such that o_p is not t -fixed, and $x'_{t,o_p} < d_{o_p}$.

(ii) Set $x'_{t,o_p} = x'_{t,o_p} + \min\{d_{o_p} - x'_{t,o_p}, \frac{W_t - \sum_{i \in \mathcal{O}_h} w_i x'_{t,i}}{w_{o_p}}\}$.

Similarly, we modify Algorithm 1 as follows:

(b') For $i \in [n]$, if i is (t, c) -fixed, set $\bar{x}_{t,i} = c$.

(c') Select the smallest $p \in \mathbb{N}$ such that o_p is not t -fixed, and $\bar{x}_{t,o_p} < d_{o_p}$.

(d') Find by binary search the biggest integer $\alpha \leq d_{o_p} - \bar{x}_{t,o_p}$ such that $\alpha w_{o_p} \leq W_t - \sum_{i \in \mathcal{O}_h} w_i \bar{x}_{t,i}$. Set $\bar{x}_{t,o_p} = \bar{x}_{t,o_p} + \alpha$. If $\alpha = d_{o_p} - \bar{x}_{t,o_p}$, go to (c'); else, stop.

Again, the juxtaposition of vectors \bar{x}_h gives the required $(1 - 3\varepsilon)$ -approximated solution \bar{x} . \square

3 Improved disjunctive relaxation for min-K

3.1 Overview of the proof technique

Let us first recall the disjunctive relaxation from [6]. The classical integer programming formulation for MIN-K can be stated as $\min\{c^T x : w^T x \geq b, x \in \{0, 1\}^n\}$, and a natural LP relaxation can be obtained by removing the integrality constraints. Since we are assuming that the objective function is fixed, we can suppose items to be sorted so that $1 = c_1 \geq c_2 \geq \dots \geq c_n$. Let $Q := \{x \in \{0, 1\}^n : w^T x \geq b\}$ be the family of all feasible solutions to MIN-K. For $j \in [n]$, let $Q_j \subseteq Q$ be the set of solutions such that $x_i = 0$ for $i < j$ and $x_j = 1$. One has $Q = \bigcup_{j \in [n]} Q_j$. It is well known that the natural LP relaxation $P_j := \{x \in [0, 1]^n : w^T x \geq b, x_i = 0 \text{ for } i < j, x_j = 1\}$ of Q_j has integrality

gap 2. Hence $\text{CONV}(\cup_{j=1}^n P_j)$ is a relaxation for $\text{CONV}(Q)$ with integrality gap 2. Bienstock and McClosky [7] provided, for each j , a relaxation for $\text{CONV}(Q_j)$ achieving integrality gap $1 + \varepsilon$. It is in fact enough to provide such a relaxation for $\text{CONV}(Q_1)$, as the others would follow by redefining $n' = n - j + 1$, items j, \dots, n as $1, \dots, n'$, and scaling costs so that $c'_1 = 1$. Their relaxation is as follows.

1. Partition set $\{2, \dots, n\}$ into the following *buckets*¹:

$$S_k := \{i \in \{2, \dots, n\} : (1 + \varepsilon)^{-k+1} \geq c_i > (1 + \varepsilon)^{-k}\}, \quad \forall k \in [C_\varepsilon]$$

and $S_\infty = \{i \in \{2, \dots, n\} : c_i \leq (1 + \varepsilon)^{-C_\varepsilon}\}$. Note that $(1 + \varepsilon)^{-C_\varepsilon} \leq \varepsilon$.

2. For $\rho \in \{0, 1, \dots, \lceil 1/\varepsilon \rceil\}^{C_\varepsilon}$ let Q_ρ be the set of all solutions in Q_1 where the number of items taken from S_k is exactly ρ_k if $\rho_k < \lceil 1/\varepsilon \rceil$, and it is at least $\lceil 1/\varepsilon \rceil$ otherwise. Again $Q_1 = \cup_\rho Q_\rho$ and each $\text{CONV}(Q_\rho)$ can be relaxed to

$$P_\rho := \{x \in P_1 : \sum_{i \in S_k} x_i = \rho_k, \forall k : \rho_k < \lceil \frac{1}{\varepsilon} \rceil \text{ and } \sum_{i \in S_k} x_i \geq \rho_k, \forall k : \rho_k = \lceil \frac{1}{\varepsilon} \rceil\}.$$

The result from [7] follows from showing that each P_ρ has integrality gap $1 + \varepsilon$. Then $\text{CONV}(\cup_\rho P_\rho)$ is a relaxation of Q_1 with integrality gap $1 + \varepsilon$.

As the first step of our relaxation, we also partition Q into the Q_j . However, the successive partition of Q_j is performed differently. Similarly to [7], our improved relaxation groups items with similar cost, but then exploits the following: a vertex x^* of a polytope like P_ρ has at most two fractional components, and they lie in the same bucket (see Lemma 12). Say those components correspond to items r and q , with $c_r \geq c_q$. It is a standard trick to round x^* to an integral solution and bound the variation of the cost as a function of c_r and c_q . If there is a non-empty bucket whose items have cost bigger than c_r , then this bucket contributes to the objective function at least as much as c_r . Hence, if there are many of those buckets, the distance between c_r and c_q can be reasonably big and still the rounding would induce a small change with respect to the total cost. We can then take (non-empty) buckets of increasing length, and still guarantee the integrality gap of $1 + \varepsilon$ (see Lemma 15). Therefore, we can partition Q_1 in a smaller number of sets, leading to a relaxation of smaller size (see Lemma 17).

3.2 The disjunctive relaxation

Because of the discussion from the previous section, in order to prove Theorem 2, we are left to provide a disjunctive relaxation for $\text{CONV}(Q_1)$. We will also assume $\varepsilon \leq 1/256$. Let $\mathcal{S} = \{S_1, \dots, S_K, S_\infty\}$ be a family of pairwise disjoint subsets of $\{2, \dots, n\}$, and $\rho \in \{1, \dots, \lceil 1/\varepsilon \rceil\}^K$. Define:

$$P(\mathcal{S}, \rho) := \{x \in \mathbb{R}^n : \begin{array}{ll} w^T x \geq b, & \\ \sum_{i \in S_k} x_i = \rho_k & \forall k \in [K] \mid \rho_k < \lceil 1/\varepsilon \rceil, \\ \sum_{i \in S_k} x_i \geq \rho_k & \forall k \in [K] \mid \rho_k = \lceil 1/\varepsilon \rceil, \\ x_1 = 1, & \\ x_i = 0 & \forall i \in \{2, \dots, n\} \setminus \cup_{k \in [K] \cup \{\infty\}} S_k, \\ 0 \leq x_i \leq 1 & \forall i \in \cup_{k \in [K] \cup \{\infty\}} S_k \end{array} \}.$$

Lemma 12. *An extreme point x^* of $P := P(\mathcal{S}, \rho)$ has at most two fractional components, and if they are two, they lie in the same bucket S_h , where $h \in [K]$.*

¹Here $C_\varepsilon = \lceil \log_{1+\varepsilon}(1/\varepsilon) \rceil$. Hence, as in Section 2, $C_\varepsilon = O(\frac{1}{\varepsilon} \log \frac{1}{\varepsilon})$. On the other hand, the bucketing is different from the one performed there.

Proof. Let x^* be an extreme point of P , and consider a set \mathcal{C} of n linearly independent constraints of P at which x^* is active. Let $\mathcal{C}' \subseteq \mathcal{C}$: basic linear algebra implies that \mathcal{C}' is also linearly independent, hence the number of variables that belong to the support of \mathcal{C}' are at least $|\mathcal{C}'|$. By Hall's Theorem, we can then find an injective map assigning to each constraint from \mathcal{C} a variables from its support. We say that the constraint is "charged" to the variable. Since $x \in \mathbb{R}^n$, the map is also surjective. Now let $0 < x_r^* < 1$. Then $r \neq 1$, i.e. $r \in S_h$ for some $h \in [k] \cup \{\infty\}$, and x_r charges either $\sum_{i \in S_k} x_i \geq \rho_k$ (or $\sum_{i \in S_k} x_i = \rho_k$), or to $w^T x \geq b$. This implies that there are at most two fractional variables per bucket, and one if $h = \infty$. Now suppose x_r^* does not charge $w^T x \geq b$: then, since the constraint it charges is tight at x^* , there exists $q \in S_h$, $q \neq r$, such that $x_r^* + x_q^* = 1$. In particular, x_q^* is fractional, and it must charge $w^T x \geq b$. Hence, we showed that each time a fractional variable does not charge $w^T x \geq b$, there is exactly one more fractional variable from the same bucket, and it charges $w^T x \geq b$. The thesis then follows from the fact that at most one variable can charge $w^T x \geq b$. \square

The lemma above gives a new insight on the extreme points of $P(\mathcal{S}, \rho)$ and it is crucial to control the decrease in the objective function when rounding. Let Γ be the set of vectors $\tau \in \mathbb{N}_0^{|\tau|}$ with the following properties:

1. $|\tau| \leq 2\sqrt{C_\varepsilon}$; 2. $\tau_k + k \leq \tau_{k+1}$ for $k \in [|\tau|]$; 3. $\tau_{|\tau|} \leq C_\varepsilon - 1$.

and for $\tau \in \Gamma$ define $\mathcal{S}(\tau)$ as follows:

- (i) For $k \in [K]$, set $S_k := \{i \in \{2, \dots, n\} : (1 + \varepsilon)^{-\tau_k} \geq c_i > (1 + \varepsilon)^{-\min\{\tau_k + k, C_\varepsilon\}}\}$.
- (ii) Set $S_\infty := \{i \in \{2, \dots, n\} : c_i < \min_{l \in S_{|\tau|}} c_l \text{ and } c_i \leq (1 + \varepsilon)^{-C_\varepsilon}\}$.

We start with some definitions and auxiliary lemmas.

Definition 13. Given $\varepsilon > 0$ and \mathcal{S} as above, define $c_{\min, k} := \min_{i \in S_k} c_i$ and $c_{\max, k} := \max_{i \in S_k} c_i$ for $k \in [K]$. We say that \mathcal{S} is (ε, c) -ordered if:

- (a) $c_{\min, k} \geq c_{\max, k+1}$ for $k \in [K - 1]$;
- (b) $\min\{c_{\min, K}, \varepsilon\} \geq \max_{i \in S_\infty} c_i$.

Lemma 14. Let \mathcal{S} be an (ε, c) -ordered partition and $\rho \in [\lceil 1/\varepsilon \rceil]^K$. An extreme point x^* of P can be rounded to an integral vector \bar{x} with cost $c(\bar{x}) \leq (1 + 2\varepsilon)c(x^*)$ if the following condition holds. Given a fractional point of x^* in bucket $h \in [K]$ one has

$$\frac{c_{\max, h}}{c_{\min, h}} \leq (1 + \varepsilon)^{d_h} \text{ for } h \in [K], \text{ with } d_h \leq \min\{h, \lceil 2\sqrt{C_\varepsilon} \rceil\}.$$

Proof. Following Lemma 12, we distinguish two cases.

Case 1: x^* has exactly one fractional component, say r . Then \bar{x} can be obtained by setting $\bar{x}_r = 1$ and $\bar{x}_i = x_i^*$ for $i \in [n] \setminus \{r\}$. x^* is clearly feasible. Moreover, $c(\bar{x}) - c(x^*) \leq c_r$. If $h = \infty$ then by (b) one has $c_r \leq \varepsilon \cdot c_1 \leq \varepsilon c(x^*)$. Otherwise, $h \in [K]$ and $\sum_{i \in S_h} x_i^*$ is fractional. Hence $\rho_h = \lceil 1/\varepsilon \rceil$, otherwise x^* would not be feasible. Then one gets

$$\frac{c_r}{c(x^*)} \leq \frac{c_{\max, h}}{c_j + \lceil 1/\varepsilon \rceil c_{\min, h}} \leq \frac{(1 + \varepsilon)^{d_h}}{\lceil 1/\varepsilon \rceil} \leq \varepsilon \frac{1 - (\varepsilon d_h)^{d_h+1}}{1 - \varepsilon d_h} \leq 2\varepsilon,$$

for ε small enough ($\leq 1/256$) using $d_h \leq \lceil 2\sqrt{C_\varepsilon} \rceil$ and

$$(1 + \varepsilon)^{d_h} = \sum_{l=0}^{d_h} \binom{d_h}{l} \varepsilon^l \leq \sum_{l=0}^{d_h} (\varepsilon d_h)^l = \frac{1 - (\varepsilon d_h)^{d_h+1}}{1 - \varepsilon d_h}.$$

Case 2: x^* has exactly two fractional entries, say r and q . From Lemma 12 and its proof, we know they are exactly in the same bucket S_h with $h \in [K]$, that $\sum_{i \in S_h} x_i^* = \rho_h \in \mathbb{Z}$, and $x_r^* + x_q^* = 1$. Assume wlog $w_r \geq w_q$. Setting $\bar{x}_r = 1$, $\bar{x}_q = 0$ and $\bar{x}_i = x_i^*$ for $i \in [n] \setminus \{r, q\}$ gives an integral feasible vector \bar{x} with the approximation guarantee:

$$\begin{aligned} \frac{c(\bar{x}) - c(x^*)}{c(x^*)} &\leq \frac{c_r - c_q}{c(x^*)} \stackrel{(*)}{\leq} \frac{c_{\max, h} - c_{\min, h}}{c_1 + (d_h - 1)c_{\max, h}} \stackrel{(\bullet)}{\leq} \frac{c_{\max, h} - c_{\min, h}}{d_h c_{\max, h}} \stackrel{(o)}{\leq} \frac{(1 + \varepsilon)^{d_h} - 1}{d_h(1 + \varepsilon)^{d_h}} \\ &\leq \frac{\frac{1 - (\varepsilon d_h)^{d_h + 1}}{1 - \varepsilon d_h} - 1}{d_h} \leq \frac{\varepsilon}{1 - \varepsilon d_h} \leq 2\varepsilon, \end{aligned}$$

where $(*)$ follows from (a) and the fact that by construction $x_1^* = 1$, and $x_i^* = 1$ for at least one $i \in S_k$, for all $k < h$, (\bullet) from $c_{\max, h} \leq c_1$, (o) from the definition of d_h , and in the last inequality we again assumed $\varepsilon \leq 1/256$ and used $d_h \leq \lceil 2\sqrt{C_\varepsilon} \rceil$. \square

Incidentally, observe that the relaxation defined in points 1.-2. from Section 3.1 is induced by an (ε, c) -ordered family, by disregarding sets S_i of the partition with $\rho_i = 0$. It also trivially satisfies the condition of Lemma 14 since $d_k = 1$, $\forall k \in [K]$, and $K \leq C_\varepsilon$. Recall that Γ is the set of vectors $\tau \in \mathbb{N}_0^{|\tau|}$ with the following properties:

1. $|\tau| \leq 2\sqrt{C_\varepsilon}$.
2. $\tau_k + k \leq \tau_{k+1}$ for $k \in [|\tau|]$.
3. $\tau_{|\tau|} \leq C_\varepsilon - 1$.

and for $\tau \in \Gamma$ we defined $\mathcal{S}(\tau)$ as follows:

- (i) For $k \in [K]$, set $S_k := \{i \in \{2, \dots, n\} : (1 + \varepsilon)^{-\tau_k} \geq c_i > (1 + \varepsilon)^{-\min\{\tau_k + k, C_\varepsilon\}}\}$.
- (ii) Set $S_\infty := \{i \in \{2, \dots, n\} : c_i < \min_{l \in S_{|\tau|}} c_l \text{ and } c_i \leq (1 + \varepsilon)^{-C_\varepsilon}\}$.

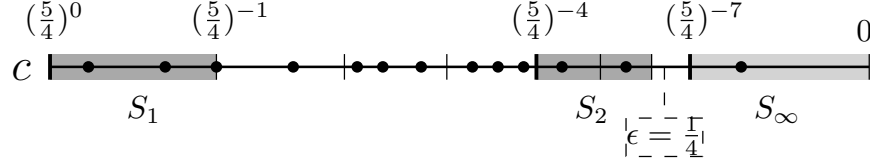


Figure 3: An example of the bucketing $\mathcal{S}(\tau)$ for $\tau = (0, 4)$ and $\varepsilon = 1/4$.

Lemma 15. *Let $\tau \in \Gamma$ and $\mathcal{S} = \mathcal{S}(\tau)$. An extreme point of $P := P(\mathcal{S}, \rho)$ can be rounded to an integral vector \bar{x} with cost $c(\bar{x}) \leq (1 + 2\varepsilon)c(x^*)$.*

Proof. It is enough to show that \mathcal{S} satisfies the conditions from Lemma 14. One immediately checks that \mathcal{S} is an (ε, c) -ordered partition. As $|\tau| \leq 2\sqrt{C_\varepsilon}$, we only need to prove that $d_h \leq h$, for each $h \in [K]$. This follows from the fact that $c_{\max, h} \leq (1 + \varepsilon)^{-\tau_h}$ and $c_{\min, h} \geq (1 + \varepsilon)^{-(\tau_h + h)}$, hence $d_h \leq -\tau_h + (\tau_h + h) = h$. \square

Lemma 16. *For any solution $\hat{x} \in Q_1$ there exist $\tau \in \Gamma$ and $\rho \in \lceil [1/\varepsilon] \rceil^{|\tau|}$ such that $\hat{x} \in P(\mathcal{S}(\tau), \rho)$.*

Proof. We iteratively construct τ as follows:

- 1) $\tau_1 = \min\{\hat{k} \in [C_\varepsilon] : \exists \hat{i} > 1 \text{ s.t. } \hat{x}_{\hat{i}} = 1, (1 + \varepsilon)^{-\hat{k}} \geq c_{\hat{i}} > (1 + \varepsilon)^{-\hat{k}-1}\}$;

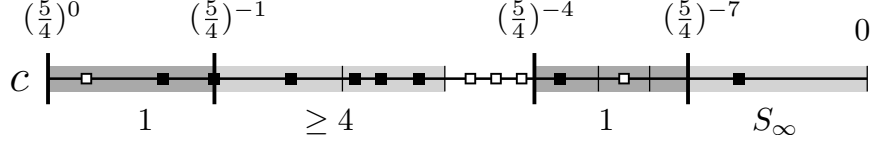


Figure 4: An example of a knapsack solution \hat{x} induced by the items marked with full-squares. The construction in Lemma 16 covers \hat{x} with $\tau = (0, 1, 4)$ and $\rho = (1, 4, 1)$ for $\varepsilon = 1/4$.

2) Given τ_k , as long as the set

$$R_{k+1} := \{\hat{k} \in [C_\varepsilon] : \hat{k} \geq \tau_k + k, \exists \hat{i} > 1 \text{ s.t. } \hat{x}_{\hat{i}} = 1, (1 + \varepsilon)^{-\hat{k}} \geq c_{\hat{i}} > (1 + \varepsilon)^{-\hat{k}-1}\}$$

is non-empty, define $\tau_{k+1} = \min\{\hat{k} \in R_{k+1}\}$.

First observe that step 2) is repeated at most $\lceil 2\sqrt{C_\varepsilon} \rceil - 1$ times, since $\sum_{k=1}^{\lceil 2\sqrt{C_\varepsilon} \rceil} k \geq C_\varepsilon$. Hence $|\tau| \leq \lceil 2\sqrt{C_\varepsilon} \rceil$. One easily concludes then that $\tau \in \Gamma$.

Now choose ρ such that $\rho_k = \min\{|\text{supp}(\hat{x}) \cap S_k|, \lceil 1/\varepsilon \rceil\}$ for $k \in [|\tau|]$. Let us verify that $\hat{x} \in P(\mathcal{S}(\tau))$. Let $i \in \{2, \dots, n\}$ such that $\hat{x}_i = 1$. All we need to show is that, if $c_i > (1 + \varepsilon)^{-C_\varepsilon}$, then $i \in S_k$ for some $k \in [|\tau|]$, since the feasibility of \hat{x} would then follow by definition of ρ . Let \hat{k} be the maximum k such that $(1 + \varepsilon)^{-\tau_k} \geq c_i$. If $c_i > (1 + \varepsilon)^{-(\tau_k + k)}$, then $i \in S_k$; else, the maximality of k is contradicted. \square

Lemma 17. *The number of possible pairs $(\mathcal{S}(\tau), \rho)$ with $\tau \in \Gamma$ and $\rho \in [\lceil \frac{1}{\varepsilon} \rceil]^{|\tau|}$ are $(\lceil 1/\varepsilon \rceil)^{O(\sqrt{C_\varepsilon})}$.*

Proof. $|\Gamma| = C_\varepsilon^{O(\sqrt{C_\varepsilon})}$, since $\tau_k \leq C_\varepsilon$ for $k \in [|\tau|]$ and $|\tau| \leq \lceil 2\sqrt{C_\varepsilon} \rceil$ by construction. Having that $\rho \in [\lceil 1/\varepsilon \rceil]^{|\tau|}$ we get the bound:

$$C_\varepsilon^{O(\sqrt{C_\varepsilon})} \cdot \left(\left\lceil \frac{1}{\varepsilon} \right\rceil\right)^{\lceil 2\sqrt{C_\varepsilon} \rceil} \leq \left[2 \frac{1}{\varepsilon} \ln \frac{1}{\varepsilon}\right]^{O(\sqrt{C_\varepsilon})} \cdot \left(\left\lceil \frac{1}{\varepsilon} \right\rceil\right)^{\lceil 2\sqrt{C_\varepsilon} \rceil} = \left(\frac{1}{\varepsilon}\right)^{O(\sqrt{C_\varepsilon})}$$

$$\text{Let } \hat{P}_1 := \text{CONV}\left(\bigcup_{\tau \in \Gamma} \bigcup_{\rho \in [\lceil \frac{1}{\varepsilon} \rceil]^{|\tau|}} P(\mathcal{S}(\tau), \rho)\right).$$

\square

We can now prove Theorem 2.

Proof of Theorem 2. Let $\hat{x} \in \hat{P}_1 \cap \{0, 1\}^n$. Hence $\hat{x} \in P(\mathcal{S}(\tau), \rho)$ for some $\tau \in \Gamma$ and $\rho \in [\lceil \frac{1}{\varepsilon} \rceil]^{|\tau|}$. Since constraints from $\text{CONV}(Q_1)$ are also valid for $P(\mathcal{S}(\tau), \rho)$, we conclude that $\hat{x} \in Q_1$. Conversely, if $\hat{x} \in Q_1$, $\hat{x} \in \hat{P}_1$ by Lemma 16. Hence \hat{P}_1 is indeed a relaxation for $\text{CONV}(Q_1)$. Since each $P(\mathcal{S}(\tau), \rho)$ has $O(n)$ variables and constraint, \hat{P}_1 can be described with a system of linear inequalities of size $n(1/\varepsilon)^{O(\sqrt{C_\varepsilon})}$ by Lemma 17. The thesis then follows from the fact that $Q = \bigcup_{j \in [n]} Q_j$ and Lemma 15.

Acknowledgements. Yuri Faenza's research was partially supported by the SNSF *Ambizione* fellowship PZ00P2_154779 *Tight formulations of 0/1 problems*.

References

- [1] Rumen Andonov, Vincent Poirriez, and Sanjay Rajopadhye. Unbounded knapsack problem: Dynamic programming revisited. *Europ. Journ. Oper. Res.*, 123(2):394 – 407, 2000.
- [2] Egon Balas. Disjunctive programs: cutting planes from logical conditions. In O.L. et al. Mangasarian, editor, *Nonlinear Programming*, volume 2, page 279–312. 1975.

- [3] Egon Balas and Pierre Bonami. Generating lift-and-project cuts from the lp simplex tableau: open source implementation and testing of new variants. *Mathematical Programming Computation*, 1(2):165–199, 2009.
- [4] Abbas Bazzi, Samuel Fiorini, Sangxia Huang, and Ola Svensson. Small extended formulation for knapsack cover inequalities from monotone circuits. In *Proc. of SODA '17*, to appear.
- [5] Pietro Belotti, Leo Liberti, Andrea Lodi, Giacomo Nannicini, and Andrea Tramontani. Disjunctive inequalities: applications and extensions. *Wiley Encycl. of Oper. Res. and Manag. Sc.*, 2011.
- [6] Daniel Bienstock. Approximate formulations for 0-1 knapsack sets. *Oper. Res. Lett.*, 36(3):317–320, May 2008.
- [7] Daniel Bienstock and Benjamin McClosky. Tightening simple mixed-integer sets with guaranteed bounds. *Math. Progr.*, 133(1):337–363, 2012.
- [8] Daniel Bienstock, Jay Sethuraman, and Chun Ye. Approximation algorithms for the incremental knapsack problem via disjunctive programming. *CoRR*, abs/1311.4563, 2013.
- [9] Maurice Cheung and David B. Shmoys. A primal-dual approximation algorithm for min-sum single-machine scheduling problems. In *Proc. of Approx-Random '01*, pages 135–146. 2011.
- [10] Szymon Dudycz and Carsten Moldenhauer. Approximated extended formulations for the knapsack cover problem. *Technical report, EPFL*, 2015.
- [11] Yuri Faenza and Laura Sanità. On the existence of compact ϵ -approximated formulations for knapsack in the original space. *Oper. Res. Lett.*, 43(3):339–342, 2015.
- [12] Lisa Fleischer, Michel X. Goemans, Vahab S. Mirrokni, and Maxim Sviridenko. Tight approximation algorithms for maximum general assignment problems. In *Proc. of SODA '06*, pages 611–620, 2006.
- [13] Jeffrey Hartline. Incremental optimization. *PhD thesis, Cornell University*, 2008.
- [14] Oscar H. Ibarra and Chul E. Kim. Fast approximation algorithms for the knapsack and sum of subset problems. *J. ACM*, 22(4):463–468, October 1975.
- [15] Klaus Jansen. A fast approximation scheme for the multiple knapsack problem. In *Proc. of SOFSEM 2012*, pages 313–324. 2012.
- [16] Richard M. Karp. Reducibility among combinatorial problems. In *Complexity of Computer Computations*, pages 85–103. Springer US, 1972.
- [17] Hans Kellerer, Ulrich Pferschy, and David Pisinger. *Knapsack problems*. Springer, 2004.
- [18] Bala Krishnamoorthy and Gábor Pataki. Column basis reduction and decomposable knapsack problems. *Discr. Opt.*, 6(3):242 – 270, 2009.
- [19] Adam Kurpisz, Samuli Leppänen, and Monaldo Mastrolilli. On the hardest problem formulations for the 0/1 lasserre hierarchy. In *Proc. of ICALP '15*, pages 872–885. 2015.
- [20] Silvano Martello and Paolo Toth. *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, Inc., New York, NY, USA, 1990.
- [21] George B. Mathews. On the partition of numbers. *Proc- of the London Math. Soc.*, s1-28(1):486–490, 1896.

- [22] Alexander Schrijver. *Combinatorial Optimization - Polyhedra and Efficiency*. Springer, 2003.
- [23] Jay Sethuraman and Chun Ye. Personal communication, 2016.
- [24] Alexa Sharp. Incremental algorithms: Solving problems in a changing world. *PhD thesis, Cornell University*, 2007.
- [25] Martin Skutella. An introduction to network flows over time. In *Research Trends in Combinatorial Optimization*, pages 451–482, 2008.
- [26] Chun Ye. On the trade-offs between modeling power and algorithmic complexity. *PhD thesis, Columbia University*, 2016.

A Notation

We refer to [22] for basic definitions and facts on approximation algorithms and polytopes. Given an integer k , we write $[k] := \{1, \dots, k\}$ and $[k]_0 := [k] \cup \{0\}$. Given a polyhedron $Q \subseteq \mathbb{R}^n$, a *relaxation* $P \subseteq \mathbb{R}^n$ is a polyhedron such that $Q \subseteq P$ and the integer points in P and Q coincide. The *size* of a polyhedron is the minimum number of facets in an extended formulation for it, which is well-known to coincide with the minimum number of inequalities in any linear description of the extended formulation.

B Background on disjunctive programming

Introduced by Balas [2] in the 70s, it is based on “covering” the set by a small number of pieces which admit an easy linear description. More formally, given a set $Q \subseteq \mathbb{Z}^n$ we first find a collection $\{Q_j\}_{j \in [m]}$ such that $Q = \bigcup_{j \in [m]} Q_j$. If there exist polyhedra $P_j, j \in [m]$ with bounded integrality gap and $P_j \cap \mathbb{Z}^n = Q_j$, then $P := \text{CONV}(\bigcup_{j \in [m]} P_j)$ is a relaxation of $\text{CONV}(Q)$ of with the same guarantee on the integrality gap. Moreover, one can describe P with (roughly) as many inequalities as the sum of the inequalities needed to describe the P_j . A variety of benchmarks of mixed integer linear programs (MILPs) have shown the improved performances of branch-and-cut algorithms by efficiently generated disjunctive cuts [3]. Branch-and-bound algorithms for solving MILP also implicitly use disjunctive programming. The branching strategy based on thin directions that come from the Lenstra’s algorithm for integer programming in fixed dimension has shown good results in practice for decomposable knapsack problems [18]. For further applications of disjunctive cuts in both linear and non-linear mixed integer settings see [5].